JOURNAL OF MULTIVARIATE ANALYSIS 12, 402–414 (1982)

Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates

GORDON J. JOHNSTON*

University of N. Carolina

Communicated by the Editors

Let (X, Y) have regression function $m(x) = E(Y \mid X = x)$, and let X have a marginal density $f_1(x)$. We consider two nonparameteric estimates of $m(x)$: the Watson estimate when f_1 is known and the Yang estimate when f_1 is known or unknown. For both estimates the asymptotic distribution of the maximal deviation from $m(x)$ is proved, thus extending results of Bickel and Rosenblatt for the estimation of density functions.

1. INTRODUCTION

Let $(X_1, Y_1),..., (X_n, Y_n)$ be a random sample from a bivariate population with distribution function $F(x, y)$ and density $f(x, y)$. Let F_1, f_1, (F_2, f_2) denote the marginal distribution and density of $X(Y)$. We are interested in estimating the unknown regression function $m(x) = E(Y \mid X = x)$ without making assumptions about either m or the distributional form of F. In this paper we consider two classes of estimates of $m(x)$. The first is due to Watson (1964) (see also Watson and Leadbetter [16], Parzen [8]); motivated by the formula

$$m(x) = \left\{\int yf(x, y)\, dy\right\} \Big/ f_1(x),$$

Received October, 1979; revised October 9, 1981.

AMS 1980 subject classification: Primary 62G05; Secondary 62J05.

Key words and phrases: Density estimates, nonparametric regression estimates, maximal deviations, Gaussian processes.

* The work of this author was supported by the Air Force Office of Scientific Research under Grants AFOSR-75-2796 and AFOSR-80-0080. The author is employed by Harris Corporation Government Information Systems Division, P.O. Box 94,000, Melbourne, Florida 32901.

0047-259X/82/030402–13$02.00/0

we define

$$m_n(x) = \left\{(n\varepsilon_n)^{-1} \sum Y_i K((x - X_i)/\varepsilon_n\right\} \left\{(n\varepsilon_n)^{-1} \sum K((x - X_i)/\varepsilon_n)\right\}^{-1} \tag{1.1}$$

and

$$\bar{m}_n(x) = \left\{(n\varepsilon_n)^{-1} \sum Y_i K((x - X_i)/\varepsilon_n)\right\} \Big/ f_1(x), \tag{1.2}$$

the latter appropriate if f_1 is known. Here $\varepsilon_n \to 0$ and K is a smooth density function symmetric about zero.

Analysis of $m_n(\cdot)$ is somewhat complicated by the fact that it is a ratio of two random variables. Yang [17] avoids this problem by defining and proving consistency of

$$M_n(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_n(X_i) - F_n(x))/\varepsilon_n), \tag{1.3}$$

$$\bar{M}_n(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_1(X_i) - F_1(x))/\varepsilon_n), \tag{1.4}$$

where F_n is the empirical distribution of $X_1,..., X_n$ and $\bar{M}_n(\cdot)$ is appropriate when F_1, f_1 are known. Briefly, Yang's estimates are motivated by consideration of statistics of the form $n^{-1} \sum J(i/(n+1))\, H(X_{(i)}, Y_{(i:n)})$, where $Y_{(i:n)}$ is the concomitant of the ith-order statistic $X_{(i)}$ (see also Yang [18]).

In the parametric normal linear regression model, (X, Y) has a bivariate normal distribution, $m(x)$ is linear in x, and one can derive uniform confidence bands for $m(x)$. In this paper, where neither F nor the form of m are known, we are able to obtain uniform confidence bands for the regression function $m(x)$. More specifically, we extend the results of Bickel and Rosenblatt [3] and Rosenblatt [11] to obtain the limit distribution of the maximal deviation

$$\sup\{|g_n(x) - m(x)|: 0 \leqslant x \leqslant 1\}, \tag{1.5}$$

where g_n is given by one of (1.2)–(1.4).

Obtaining the limit distribution of (1.5) when using the special estimates (1.2) and (1.4) (special because they require f_1 known) is a conceptually simple extension of Rosenblatt's [11] results. However, our real interest is in the useful estimate (1.3), for which we obtain the limit distribution of (1.5) by showing that $M_n(x) - \bar{M}_n(x)$ is uniformly sufficiently close to zero. We have been unable to obtain useful results for (1.1), the major technical difficulty being its form as a ratio of two random variables.

Related Literature

Schuster [12] and Johnston [5] give different conditions for the pointwise asymptotic normality of (1.1) and (1.2). Schuster and Yakowitz [13] give rates of almost sure convergence to zero for the maximal deviation (1.5) using (1.1). Priestly and Chao [9] and Benedetti [1] consider an estimate closely related to (1.2) for the case that X is nonstochastic. Stone [14] and Lai [6] give weak conditions for consistency of nearest-neighbor estimates. The work of Marcondes [7] is also of interest.

2. Assumptions and a Preliminary Result

Define $m_n^*(x) = f_1(x)\,\bar{m}_n(x)$ and $s(x) = E(Y^2 \mid X = x)$. In this section we prove maximal deviation results for $m_n^*(\cdot)$, applying these results to (1.2)–(1.4) in the next section. Let $\{a_n\}$ be a sequence of constants with $a_n \to \infty$ as $n \to \infty$.

Assumptions

(A1) For all n and some $c < \infty$,

$$(\log n)\,\varepsilon_n^{-3} \int_{|y| \geqslant a_n} y^2 f_2(y)\,dy \leqslant c.$$

(A2) $a_n \varepsilon_n^{-1/2} n^{-1/6} (\log n)^2 \to 0$ as $n \to \infty$, and $(\log n)^{-1}\,(n\varepsilon_n^{1/2}) \to \infty$.

(A3) $(\log n) \sup_{0 \leqslant x \leqslant 1} \int_{|y| \geqslant a_n} y^2 f(x, y)\,dy \to \infty$ as $n \to \infty$.

(A4) There exists $\eta > 0$ such that $0 \leqslant x \leqslant 1$ and $n \geqslant 1$ implies

$$|g_n(x)| = \left| \int_{-a_n}^{a_n} y^2 f(x, y)\,dy \right| > \eta.$$

(A5) The kernel function K vanishes outside a finite interval $[-A, A]$ and is absolutely continuous on $[-A, A]$, $A > 1$.

(A6) The marginal density $f_1(x)$ is continuous and positive on an open interval containing $[0, 1]$.

(A7) For g_n defined by (A4), $\{g_n^{1/2}\}$ have uniformly bounded and continuous first derivatives on $[-A, A]$.

(A8) Both $f(x)\,s(x)$ and $E(|Y| \mid X = x)\,f(x)$ are bounded for $0 \leqslant x \leqslant 1$.

Note that (A1) and (A3) hold if Y is bounded and $a_n = \log\log n$, while

(A1), (A3) and (A4) hold with $a_n = n^\beta$ ($\beta > 0$, β near zero) if (X, Y) are jointly normally distributed.

THEOREM 1. *Suppose* (A1)–(A8) *hold and* $\varepsilon_n = n^{-\delta}$ $(0 < \delta < \frac{1}{3})$. *Define*

$$Y_n(t) = (n\varepsilon_n)^{1/2}\,(m_n^*(t) - Em_n^*(t))(s(t)\,f(t))^{-1/2}.$$

Then

$$\{P(2\delta \log n)^{1/2}\,[\sup_{0 \leqslant t \leqslant 1} |Y_n(t)|/(\lambda(K))^{1/2} - d_n] < x\} \to e^{-2e^{-x}}, \qquad (2.1)$$

where $\lambda(K) = \int K^2(u)\,du$ *and*

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2}\left\{\log\left(\frac{c_1(K)}{\pi^{1/2}}\right) + \frac{1}{2}[\log \delta + \log\log n]\right\}$$

if $c_1(K) = (K^2(A) + K^2(-A))(2\lambda(K))^{-1} > 0$ *and otherwise*

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log\left(\frac{c_2(K)}{2\pi}\right),$$

where

$$c_2(K) = (2\lambda(K))^{-1}\int [K'(u)]^2\,du.$$

The similarity of Theorem 2.1 to the main results of Bickel and Rosenblatt [3] and Rosenblatt [11] is obvious. The major technical difficulty in adapting their proofs for density estimates is the possible unboundedness of Y, which is the reason for the somewhat awkward form of (A1)–(A4). The proof of Theorem 2.1, which is given in Appendix A, closely follows Rosenblatt's [11] argument.

In applications, we would want to replace $Em_n^*(t)$ (in the definition of Y_n) by $m(t)$; this results in the following corollary.

COROLLARY 2.1. *Suppose in Theorem* 2.1 *that* $\frac{1}{5} < \delta < \frac{1}{3}$, *that* $\int u^2 K(u)\,du < \infty$ *and that* $m(t)\,f(t)$ *has two bounded continuous derivatives. Then* (2.1) *holds for the process*

$$Y_n^*(t) = (n\varepsilon_n)^{1/2}\,[m_n^*(t) - m(t)\,f(t)](s(t)\,f(t))^{-1/2}.$$

Remark. While all results are stated for suprema over the interval $[0, 1]$, they extend to arbitrary finite intervals $[a, b]$ with no change except that (A4), (A6) and (A8) must hold for $a \leqslant x \leqslant b$, and $A > \max(|a|, |b|)$.

3. Applications to (1.2)–(1.4)

The limiting distribution of the maximal deviation of (1.2) is particularly simple since

$$\begin{aligned} Y_n(t) &= (n\varepsilon_n)^{1/2}\,(\bar{m}_n(t) - E\bar{m}_n(t))(f(t)/s(t))^{1/2}, \\ Y_n^*(t) &= (n\varepsilon_n)^{1/2}\,(\bar{m}_n(t) - m(x))(f(t)/s(t))^{1/2}. \end{aligned} \tag{3.1}$$

The distribution for (1.4) is also fairly simple to derive from Theorem 2.1. One notes that if (A6) is strengthened as in Rosenblatt (1976) to

(A6′) $f_1(x)$ is continuous and positive on the smallest interval containing its support,

then $Z_i = F_1(X_i)$ is uniformly distributed, $E(Y \mid Z = F(u)) = m(u)$, $E(Y^2 \mid Z = F(u)) = s(u)$ and $f_Z(u) = 1$. Thus, Theorem 2.1 and Corollary 2.1 will hold for the processes

$$\begin{aligned} Y_{n1}(t) &= (n\varepsilon_n)^{1/2}\,[\bar{M}_n(t) - E\bar{M}_n(t)]\,s(t)^{-1/2}, \\ Y_{n2}(t) &= (n\varepsilon_n)^{1/2}\,[\bar{M}_n(t) - m(t)]\,s(t)^{-1/2}. \end{aligned} \tag{3.2}$$

Finally, we consider (1.3), which is applicable in the usual case that the marginal density $f_1(X)$ of X is unknown. Consider the following assumptions.

(B1) $E\,|Y| < \infty$.

(B2) $E(Y \mid X = F^{-1}(u)) = g(u)$ has two bounded derivatives on $[0, 1]$.

(B3) $E(|Y| \mid X = F^{-1}(u)) = h(u)$ is bounded on $[0, 1]$.

(B4) There exists $a_n \to \infty$ with $a_n^2 \log n / n\varepsilon_n^3 \to 0$ and

$$n^{1/2} \int_{|y| \geqslant a_n} |y|\, dF_2(y) \to 0 \qquad \text{as} \quad n \to \infty.$$

(B5) K has three continuous bounded derivatives on its support.

Theorem 3.1. *Assume* (A1)–(A8), (A6′), (B1)–(B5). *Then if* $0 < F(a) < F(b) < 1$,

$$(n\varepsilon_n \log n)^{1/2} \sup_{a \leqslant u \leqslant b} |M_n(u) - \bar{M}_n(u)| \xrightarrow{P} 0,$$

so that Theorem 2.1 *and Corollary* 2.1 *hold for the processes defined by substituting* M_n *for* $\bar{M}_n$ *in* (3.2) (*the proof is given in Appendix* B).

Theorem 3.1 *can be used to construct uniform confidence intervals for the regression function as follows.*

COROLLARY 3.1. *Assuming Theorem 3.1 holds, an approximate* $(1-\alpha)\times 100\%$ *confidence band over an interval* $[a,b]$ *is*

$$M_n(u) \pm (n\varepsilon_n)^{-1/2}\,[s(u)\,\lambda(K)]^{1/2}\,[d_n + c(\alpha)(2\delta\log n)^{-1/2}],$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ *(for practical applications, one would estimate* $s(u)$*).*

APPENDIX A

We begin with two lemmas. Let W be Brownian motion on $(-\infty,\infty)$ and let K be a symmetric density which satisfies (A5).

LEMMA A.1 (Bickel and Rosenblatt [3]). *Let* d_n *and* $\lambda(K)$ *be as in Theorem 2.1 and let* $\varepsilon_n = n^{-\delta}(0<\delta<\frac{1}{2})$. *Define*

$$V_n(t) = \varepsilon_n^{-1/2}\int K((t-x)/\varepsilon_n)\,dW(x).$$

Then

$$P\{(2\delta\log n)^{1/2}\{\sup_{0\leqslant t\leqslant 1}|V_n(t)|/(\lambda(K))^{1/2} - d_n\} < x\} \to e^{-2e^{-x}}.$$

LEMMA A.2 (Revesz [10], Rosenblatt [11]). *Let* $(\mathbf{X}_1,\dots,\mathbf{X}_n,\dots)$ *be independent and uniformly distributed on* $[0,1]^2$. *One can construct a sequence of Brownian bridges* B_n *such that*

$$\sup\left\{\left|n^{1/2}\left(F_n(\mathbf{x}) - \prod_{j=1}^{2} x_j\right) - B_n(\mathbf{x})\right|\right\} = O(n^{-1/6}(\log n)^{3/2}) \qquad \text{a.s.},$$

where $\mathbf{x} = (x_1, x_2)$ *and* sup *is over the set* $0\leqslant x_1,\ x_2\leqslant 1$.

When Y *is bounded, since* K *vanishes off an interval, the proof of Theorem 2.1 is an easy extension of Rosenblatt's [11] result; the relevant change of variables formula is*

$$\begin{aligned}\int_{-A}^{A}\int_{-B}^{B} f(x,y)\,dg(x,y) &= \int_{-A}^{A}\int_{-B}^{B} g(x,y)\,df(x,y)\\ &\quad + \int_{-A}^{A} f(B,y)\,dg(B,y) - \int_{-A}^{A} f(-B,y)\,dg(B,y)\\ &\quad + \int_{-B}^{B} g(x,A)\,df(x,A) - \int_{-B}^{B} g(x,-A)\,df(x,-A).\end{aligned} \tag{A.1.1}$$

Hence, for the case Y unbounded, we merely sketch the proof, pointing out where the various assumptions are used. Let $Z_n(x, y) = n^{1/2}(F_n(x, y) - F(x, y))$, so that

$$Y_n(t) = [s(t) f(t)]^{-1/2} \varepsilon_n^{-1/2} \iint yK((t-x)/\varepsilon_n)\, dZ_n(x, y). \tag{A.1.2}$$

We also make the definition

$$Y_{0,n}(t) = [s(t) f(t)]^{-1/2} \varepsilon_n^{-1/2} \iint_{|y| \leqslant a_n} yK((t-x)/\varepsilon_n)\, dZ_n(x, y). \tag{A.1.3}$$

Let $\|V(\cdot)\| = \sup\{|V(t)|: 0 \leqslant t \leqslant 1\}$.

LEMMA A.3. $\|Y_n - Y_{0n}\| = o_p((\log n)^{-1/2})$.

Proof. $\|Y_n - Y_{0n}\| \leqslant \varepsilon_n^{-1/2} \|g^{-1/2}\| \, \|U_n\|$, where $g(x) = f(x)\, s(x) = \int y^2 f(x, y)\, dy$ and

$$U_n(x) = \iint_{|y| > a_n} yK((t-x)/\varepsilon_n)\, dZ_n(x, y).$$

By (A4), $\|g^{-1/2}\| > 0$. It is easy to show by Markov's inequality and (A1) that $U_n(x) \to^p 0$ for any $0 \leqslant x \leqslant 1$. The lemma will follow if U_n is tight on $D[0, 1]$. By (A5) and the Schwarz inequality,

$$\begin{aligned} &E\,|U_n(t) - U_n(t_1)|\,|U_n(t_2) - U_n(t)| \\ &\quad \leqslant M_0(\log n)\, \varepsilon_n^{-3}\, |t_1 - t|\, |t_2 - t| \int_{|y| > a_n} y^2 f_2(y)\, dy, \end{aligned}$$

verifying tightness by (A1) and Theorem 15.6 of Billingsley [4]. ∎

Define

$$\begin{aligned} s_n(t) &= E\{Y^2 I(|Y| \leqslant a_n) \mid X = t\}, \\ Y_{1n}(t) &= (s_n(t)/s(t))^{-1/2}\, Y_{0n}(t). \end{aligned} \tag{A.1.4}$$

Our next approximation is

LEMMA A.4. $\|Y_{0n} - Y_{1n}\| = o_p((\log n)^{-1/2})$.

Proof. We will later prove that

$$(\log n)^{1/2}\, \{\|Y_{1n}\|\, \{\lambda(K)\}^{-1/2} - d_n\}$$

has a limit distribution. Since $d_n = O((\log n)^{1/2})$, this means $\|Y_{1n}\| =$

$O_p((\log n)^{1/2})$. By (A3), (A7) and (A6), recalling that $g_n(x) = f(x)\, s_n(x)$, we have

$$\|(s_n/s)^{-1/2} - 1\| = o((\log n)),$$

completing the proof. ∎

Next let T be the transformation of (X, Y) to a uniform random variable on $[0, 1]^2$ ((26), (27) of Rosenblatt [11]). Define

$$Y_{2n}(t) = [s_n(t)\, f(t)]^{-1/2}\, \varepsilon_n^{-1/2} \iint_{|y| \leqslant a_n} yK((t-x)/\varepsilon_n)\, dB_n(T(x, y)),$$

$$Y_{3n}(t) = [s_n(t)\, f(t)]^{-1/2}\, \varepsilon_n^{-1/2} \iint_{|y| \leqslant a_n} yK((t-x)/\varepsilon_n)\, dW_n(T(x, y)),$$

where $B_n(u, s) = W_n(u, s) - usW_n(1, 1)$ (W_n here is the two-dimensional Wiener process).

LEMMA A.5.

$$\|Y_{1n} - Y_{2n}\| = O_p(a_n \varepsilon_n^{-1/2} n^{-1/6} (\log n)^{3/2}) = o_p((\log n)^{-1/2}) \qquad (by\ (\text{A2}))$$

and

$$\|Y_{2n} - Y_{3n}\| = o_p((\log n)^{-1/2}).$$

Proof. Using Lemma A.2, (A5) and the integration by parts formula (A.1.1), extremely detailed calculations show

$$\begin{aligned}\varepsilon_n^{1/2}\, \|g_n\|^{1/2}\, \|Y_{1n} - Y_{2n}\| &= O_p(n^{-1/6}(\log n)^{3/2}) \\ &\quad \times \left\{4a_n \int_{-A}^{A} |K'(u)|\, du + 4a_n[K(A) + K(-A)]\right\} \\ &= O_p(a_n n^{-1/6} (\log n)^{3/2}),\end{aligned}$$

completing the first part of the proof. Since the Jacobian of the transform T is $f(x, y)$, we have

$$\begin{aligned}&|Y_{2n}(t) - Y_{3n}(t)| \\ &\quad = \left|(g_n(t))^{-1/2}\, \varepsilon_n^{-1/2} \iint_{|y| < a_n} yK((t-x)/\varepsilon_n)\, f(x, y)\, dx\, dy\right| \cdot |W_n(1, 1)|.\end{aligned}$$

Thus,

$$\|Y_{2n} - Y_{3n}\| \leqslant |W_n(1, 1)| \, \|g_n^{-1/2}\| \, \varepsilon_n^{-1/2} \times \left\| \int |y| f(x, y) \, dy \, K((t-x)/\varepsilon_n) \, dx \right\|.$$

By (A8) and (A4), $\|Y_{2n} - Y_{3n}\| = O_p(\varepsilon_n^{1/2})$, completing the proof. ∎

Now define

$$Y_{4n}(t) = [s_n(t) f(t)]^{-1/2} \varepsilon_n^{-1/2} \int [s_n(x) f(x)]^{1/2} K((t-x)/\varepsilon_n) \, dW(x),$$

$$Y_{5n}(t) = \varepsilon_n^{-1/2} \int K((t-x)/\varepsilon_n) \, dW(x).$$

Since Y_{3n} and Y_{4n} are Gaussian with the same covariance function, they have the same distribution. Thus, by Lemmas A.1, A.3, A.4 and A.5, we need merely prove

LEMMA A.6. $\|Y_{4n} - Y_{5n}\| = o_p((\log n)^{1/2})$.

Proof. First note that

$$|Y_{4n}(t) - Y_{5n}(t)| = \varepsilon_n^{-1/2} \left| \int_{-A}^{A} \{(g_n(t-u\varepsilon_n)/g_n(t))^{1/2} - 1\} K(u) \, dW(t-u\varepsilon_n) \right|.$$

Since by (A7)

$$\varepsilon_n^{-1/2} \sup_{0 \leqslant t \leqslant 1} |(g_n(t \pm A\varepsilon_n)/g_n(t))^{1/2} - 1| = O(1),$$

using integration by parts and the assumptions that $g_n^{1/2}$ and K are absolutely continuous, we obtain

$$\begin{aligned} |Y_n(t) - Y_{5n}(t)| &\leqslant \varepsilon_n^{-1/2} \left| \int_{-A}^{A} W(t-u\varepsilon_n) \frac{\partial}{\partial u} \right. \\ &\qquad \times \{\{[g_n(t-u\varepsilon_n)/g_n(t)]^{1/2} - 1\} K(u)\} \, du + O_p(\varepsilon_n^{1/2}) \\ &= J_n(t) + O_p(\varepsilon_n^{1/2}). \end{aligned}$$

Note that $\varepsilon_n^{-1} \partial\{(g_n(t-u\varepsilon_n)/g_n(t))^{1/2} - 1\}/\partial u$ is uniformly bounded by (A4) and (A7), so that

$$\varepsilon_n^{-1/2} J_n(t) \leqslant \varepsilon_n^{-1} \left| \int_{-A}^{A} W(t-u\varepsilon_n)\, K'(u)[\{ g_n(t-u\varepsilon_n)/g_n(t)\}^{1/2} - 1]\, du \right.$$
$$+ C_1 \int_{-A}^{A} |W(t-u\varepsilon_n)|\, du$$
$$\leqslant C_2 \int_{-A}^{A} |W(t-u\varepsilon_n)|\, uK'(u)\, du + C_1 \int_{-A}^{A} |W(t-u\varepsilon_n)|\, du;$$

hence $\varepsilon_n^{-1/2} \|J_n\| = O_p(1)$ and $\| Y_{4n} - Y_{5n}\| = O_p(\varepsilon_n^{1/2})$, which completes the proof. ∎

Appendix B

Define

$$M_n^*(x) = (n\varepsilon_n)^{-1} \sum_{i=1}^{n} Y_i K((F_n(X_i) - F(x))/\varepsilon_n).$$

We will prove Theorem 3.1 by showing

$$(n\varepsilon_n \log n)^{1/2} \sup_{a \leqslant u \leqslant b} |M_n(u) - M_n^*(u)| \xrightarrow{P} 0 \tag{B.1.1}$$

and

$$(n\varepsilon_n \log n)^{1/2} \sup_{a \leqslant u \leqslant b} |M_n^*(u) - \bar{M}_n(u)| \xrightarrow{P} 0. \tag{B.1.2}$$

We only prove (B.1.1) as (B.1.2) is similar.

The following lemma is very similar to Lemma 1 of Bhattacharyya (1976).

Lemma B.1. *Assume that $g(u) = E[Y \mid X = F^{-1}(u)]$ has r continuous derivatives on $[0, 1]$, $r > 0$, and that K has bounded support and r bounded derivatives on the support. Then for a, b such that $0 < F(a) < F(b) < 1$,*

$$\left| \varepsilon_n^{-(r+1)} \iint yK^{(r)}((F(x) - F(z))/\varepsilon_n)\, dF(x, y) \right| = O(1)$$

uniformly in $z \in [a, b]$ as $n \to \infty$. ∎

Letting $Z_n(x, y) = F_n(x, y) - F(x, y)$, we see

$$M_n^*(u) - M(u) = \varepsilon_n^{-1} \iint y[K((F_n(x) - F(u))/\varepsilon_n) - K((F_n(x) - F_n(u))/\varepsilon_n)] \times [dZ_n(x, y) + dF(x, y)] = J_1 + J_2.$$

We first show $(n\varepsilon_n \log n)^{1/2} |J_2| \to^p 0$. Let $\xi_n(u) = (F_n(u) - F(u))/\varepsilon_n$. By (B5),

$$\begin{aligned} J_2 &= \varepsilon_n^{-1}\xi_n(u) \iint y[K'(\xi_n(u)) + \tfrac{1}{2}\xi_n(u) K''(\xi_n(u)) \\ &\quad + \tfrac{1}{6}\xi_n(u)^2 K'''(\xi_n(u) + w_n(u)/\varepsilon_n)]\, dF(x, y) \\ &= J_2^{(1)} + J_2^{(2)} + J_2^{(3)} \end{aligned}$$

where $w_n(u)$ is between $F_n(u)$ and $F(u)$. Recall that K has three bounded continuous derivatives on a compact support. This together with the fact that $\sup |F_n(x) - F(x)| = O_p(n^{-1/2})$ yields by a Taylor expansion

$$\begin{aligned} &(n\varepsilon_n \log n)^{1/2} \sup\{|J_2^{(1)}| : a \leqslant u \leqslant b\} \\ &\quad \leqslant n\varepsilon_n \log n)^{1/2} \varepsilon_n^{-2} O_p(n^{-1/2}) \\ &\quad \times \sup_u \left\{ \left| \iint yK'((F(x) - F(u))/\varepsilon_n)\, dF(x, y) \right. \right. \\ &\quad \left. + \varepsilon_n^{-1} O_p(n^{-1/2}) \int_0^1 h(t) |K''((t - F(u))/\varepsilon_n)\, dt + \varepsilon_n^{-2} O_p(n^{-1}) E\, |Y| \right\}. \end{aligned} \tag{B.1.3}$$

Applying Lemma B.1 shows that the first term on the right of (B.1.3) converges in probability to zero. Making a change of variable shows that the second term is $(n\varepsilon_n \log n)^{1/2} \varepsilon_n^{-2} O_p(n^{-1}) = o_p(1)$ by (B4). The third term is $(\log n)^{1/2} \varepsilon_n^{-3/2} O_p(n^{-1}) = o_p(1)$, also by (B4). Similar calculations apply to $J_2^{(2)}$ and $J_2^{(3)}$, so we have shown $(n\varepsilon_n \log n)^{1/} \sup\{|J_2| : a \leqslant u \leqslant b\} \to^p 0$. We thus need only prove

$$(n\varepsilon_n \log n)^{1/2} \sup\{|J_1| : a \leqslant u \leqslant b\} \xrightarrow{P} 0. \tag{B.1.4}$$

Rewrite

$$\begin{aligned} J_1(u) &= \varepsilon_n^{-1} \left[\int_{|y| > a_n} + \int_{|y| < a_n} \right] (yG_n(x, u)\, Z_n(dx, dy)) \\ &= J_1^{(1)} + J_1^{(2)}, \end{aligned}$$

where

$$G_n(x, u) = K((F_n(x) - F(u))/\varepsilon_n) - K((F_n(x) - F_n(u))/\varepsilon_n).$$

Define $Q_n(y) = F_n(y) - F(y)$ and use integration by parts to show

$$J_1^{(2)} = \varepsilon_n^{-1} \iint_{|y| < a_n} Z_n(x, y)\, dy\, G_n(dx, u) + \varepsilon_n^{-1} \int_{-a_n}^{a_n} yG_n(\infty, u)\, dQ_n(y)$$

$$+ a_n \varepsilon_n^{-1} \int \{Z_n(x, a_n) + Z_n(x, -a_n)\}\, G_n(dx, u)$$

$$= (I_1 + I_2 + I_3 + I_4)(u).$$

Now, by the mean value theorem and the boundedness of K',

$$|I_2(u)| \leqslant \varepsilon_n^{-2} O_p(n^{-1/2}) \int_{-a_n}^{a_n} |y|\, dQ_n(y).$$

By Markov's inequality, $\int_{-a_n}^{a_n} |y|\, dQ_n(y) = O_p(a_n n^{-1/2})$, whence

$$(n\varepsilon_n \log n)^{1/2} \sup_u |I_2(u)| = O_p(a_n (n\varepsilon_n^2)^{-1}) = o_p(1).$$

We deal only with $I_3(u)$, as $I_4(u)$ is similar. If $V[\cdot]$ denotes total variation,

$$\begin{aligned} |I_3(u)| &\leqslant a_n \varepsilon_n^{-1} O_p(n^{-1/2})\, V[G_n(\cdot, u)] \\ &= a_n \varepsilon_n^{-1} O_p(n^{-1/2}) \{\varepsilon_n^{-1} O_p(n^{-1/2})\} \end{aligned}$$

uniformly in $a \leqslant u \leqslant b$. Thus, by (B4), $(n\varepsilon_n \log n)^{1/2} \sup_u \{|I_3(u)| + |I_4(u)|\} \to^p 0$. Similarly,

$$\begin{aligned} &(n\varepsilon_n \log n)^{1/2} \sup\{|I_1(u)| : a \leqslant u \leqslant b\} \\ &\leqslant (n\varepsilon_n \log n)^{1/2}\, \varepsilon_n^{-1} O_p(n^{-1/2})\, V[\, G_n(x, u)] \\ &= (n\varepsilon_n \log n)^{1/2}\, \varepsilon_n^{-1} O_p(n^{-1/2})\, a_n \varepsilon_n^{-1} O_p(n^{-1/2}) \\ &= o_p(1), \end{aligned}$$

where here V denotes total variation in (x, y) over $R \times [-a_n, a_n]$.

Thus to verify (B.1.4) we must show $(n\varepsilon_n \log n)^{1/2} \sup |J_1^{(1)}| \to^P 0$. Routine calculations show

$$\varepsilon_n |J_1^{(1)}| \leqslant \varepsilon_n^{-1} O_p(n^{-1/2}) \int_{|y| > a_n} |y|\, dF_n(y) + a_n \varepsilon_n^{-1} O_p(n^{-1/2}),$$

completing the proof by (B4).

ACKNOWLEDGMENTS

The author thanks R. J. Carroll for suggesting the problem and M. R. Leadbetter for many helpful conversations.

REFERENCES

[1] BENEDETTI, J. K. (1974). *Kernel Estimation of Regression Functions.* Ph. D. Dissertation, Department of Biostatistics, University of Washington.

[2] BHATTACHARYA, P. K. (1976). An invariance principle in regression analysis. *Ann. Statist.* **4**, 621–624.

[3] BICKEL, P. J. AND ROSENBLATT, M. (1973). On some global measures of the deviation of density function estimators. *Ann. Math. Statist.* **1** 1071–1095.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures* Wiley, New York.

[5] JOHNSTON, G. (1979). Smooth nonparametric regression analysis. Institute of Statistics Mimeo Series 1253, University of North Carolina.

[6] LAI, S. L. (1977). *Large Sample Properties of K-Nearest Neighbor Procedures.* Ph. D. Dissertation, Department of Mathematics, University of California, Los Angeles.

[7] MARCONDES, N. P. (1978). *Estimation of Multivariate Densities, Conditional Distributions and Related Functions.* Ph. D. Dissertation, Department of Statistics, University of California, Berkeley.

[8] PARZEN, E. (1962). On estimation of a probability density function. *Ann. Math. Statist.* **33** 1065–1076.

[9] PRIESTLY, M. B. AND CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392.

[10] REVESZ, P. (1976) On strong approximation of the multidimensional empirical process. *Ann. Prob.* **4** 729–743.

[11] ROSENBLATT, M. (1976). On the maximal deviation of a k-dimensional density estimator. *Ann. Probab.* **4** 1009–1015.

[12] SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

[13] SCHUSTER, E. F. AND YAKOWITZ, S. (1979). Contributions to the theory of nonparametric regression, with applications to system identification. *Ann. Statist.* **7** 139–149.

[14] STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

[15] WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā, Ser. A Math. Sci.* **26** 359–372.

[16] WATSON, G. S. AND LEADBETTER, M. R., (1963). On the estimation of the probability density, I. *Ann. Math. Statist.* **34** 480–491.

[17] YANG, S. S. (1977a). Linear functions of concomitants of order statistics. Technical Report No. 7, Department of Mathematics, M.I.T.

[18] YANG, S. S. (1977b). General distribution theory of the concomitants of order statistics. *Ann. Statist.* **6** 996–1002.